

3

Paradata ‘In the Wild’

How and Where Paradata Emerges in Research Documentation

Olle Sköld and Lisa Andersson

3.1 Introduction

All manners of data-driven activities and processes, including analytical, curatorial and scholarly work, require access to data that can be used for the task at hand. Access to relevant and usable data is, however, not a primordial state of affairs in the sense that it is without preconditions. Instead, it ultimately rests on the user’s ability to determine what data looks like, how it can be accessed, and whether it is task-relevant or not. The literature describing data and work with data in different settings offers an abundance of examples. Knorr-Cetina (1999) and Berg (1996) show how physicians have to sift through large amounts of signs, tactile and visual information, and outputs from medical machinery, physical examinations, as well as the patient’s responses to clinical background questions to identify the data relevant to diagnose the medical condition at hand. In other settings, where different goals and methods are at play, similar situations emerge: Latour (1999) depicts how a group of Earth scientists arrives at a field site located at the crossroads of a savanna and the Amazonian rainforest and – by engaging in observation, indexing, cataloguing and sample-taking – transforms the vast complex of potential data offered at the site, manifested in tree varieties, different types of bushes, plants and soil types into a scholarly dataset pertinent to the question the expedition set out to answer: Is the rainforest advancing or retreating? Previous chapters in this volume have established paradata as a heterogeneous phenomenon with a broad range of examples and conceptual constructions and use cases, including facilitating secondary use of research data by offering insights into the practices and processes of data creation, curation and use (Huvila, 2022; Sköld et al., 2022). For paradata to become useful in supporting

data reuse applications, however, it (like data) must first be identified and accessed, and just as in working out what vegetation and soil compositions are relevant for understanding the dynamics of Amazonian ecosystems, the identification of paradata is a context-sensitive task that is not particularly straightforward in the sense that what paradata is relevant and important will notably vary from use case to use case.

It has been suggested that in research-based data making one of the most valuable resources for identifying paradata is research documentation, that is, the many documents, datasets, notes, communications, drafts and documentary fragments that are produced during the course of scholarly work and stored in data repositories, email inboxes, hard-drive folders, and office shelves and cupboards (Borgman, 2012; Frank et al., 2015; Gant and Reilly, 2018; Huggett, 2012; Weber et al., 2012). Two approaches to identifying paradata in research documentation can be identified in the literature that this chapter will draw on. The first approach (paradata ‘as thing’; cf. Buckland, 1991) is based on the idea that paradata can be documented, stored, managed, retrieved and used, and entails looking for descriptors and identifiers in research documentation indicating the existence of paradata (Faniel and Yakel, 2017; Huvila et al., 2022a; Huvila and Sköld, 2023). These descriptors and identifiers can be of different kinds. Nominally universal and domain-agnostic paradata identifiers exist in standards, schemes, guidelines, conceptual models and other instructions of how to describe and systematise data and, of course, in the corresponding datasets normalised by implementing these instructions (Börjesson et al., 2020; Mayernik, 2016; Zimmerman, 2008). For example, The London Charter (The London Charter Organization, 2009) and CIDOC CRM (Doerr et al., 2007) are a guideline and a standard regulating the description of archaeological visualisations and cultural heritage resources, respectively. They suggest that decisions and implicit and explicit reasoning underpinning them should be documented along with actions and events important to understanding the provenance of the item described. Paradata identifiers and descriptors are not solely found in highly structured data and data descriptions schemas, however, but are also present in more loosely organised datasets that are not systematic, but systematised to the extent required to enable sufficient analysis in their settings of origin (Baker and Yarmey, 2009; Börjesson et al., 2022b; Gitelman, 2013).

The second approach to identifying paradata suggested by earlier studies (paradata ‘as practice’) is to track paradata use cases and to learn from the gaze of paradata users where paradata emerges and how it is mobilised into action (Latour and Woolgar, 1979; Pasquetto et al., 2019; Shankar, 2009). This way of locating paradata emphasises domain knowledge and contextual insights on

a micro-meso-macro gradient (research projects, research data collaborations, disciplinary epistemic horizons) as essential resources in identifying paradata. The approach draws on a longer line of research on scholarly data that stresses that while (para)data certainly is tangible and exists within a framework of material dependencies, it is also a malleable phenomenon that manifests itself differently across use settings and disciplinary contexts depending on what the specificities of the task requiring (para)data are (Borgman, 2015; Pinch and Bijker, 1984; Wallis et al., 2013).

Drawing on insights in tracing paradata in research documentation, the aim of this chapter is to show how and where paradata emerges in data documentation and what this paradata might look like. The chapter's approach to identifying paradata draws on and intertwines the two approaches outlined above. The analysis of how paradata is identified is informed by a comprehensive interview study done within the auspices of the CAPTURE project of how archaeologists and archaeological research data professionals locate, access and use paradata for reuse purposes (i.e., building on the paradata as practice perspective). One of the main objectives of the chapter is to facilitate more tangible understandings of what paradata (as a thing) is and how it can be located. Further, this chapter assumes that paradata can be present in documentation such as notes and log files, but also in oral briefings and other interactions between colleagues. Chapter 1 outlined the scope and rationales underpinning how the present book engages with paradata and Chapter 2 provided a deep-dive into the concept of paradata. This chapter offers an overview of practical examples of paradata that anticipates both the walk-through of formalised methods for identifying, capturing and documenting paradata offered in Chapters 4 and 5, and the approaches to managing paradata discussed in Chapter 6.

The chapter is organised as follows. First, literature on three topics is reviewed to create a framework for discussing how and where paradata emerges in research documentation. The topics reviewed are foundational perspectives on data, contextual perspectives on data, and how science and scholarly work can be understood from the viewpoints of documents and documentation. After that, a practice-led study of paradata in research documentation is presented on the basis of the CAPTURE interview study (Börjesson and Sköld, 2021; the study has been previously and differently reported in e.g., Börjesson, 2021 and Börjesson et al., 2022a) conducted in 2020–2021 and comprising thirty-one interviews.¹ Table 3.2 presents the key

¹ The semi-structured interviews were held in Swedish and English, and all translations have been done by the chapter authors.

findings of the chapter and describes what research documentation can be consulted to identify and extract paradata. The table also describes key characteristics of the paradata sources and the access affordances that impact how they can be located and used.

3.2 Understanding (Para)data

Scholarship on data can provide important resources for starting to think about what paradata might be and how it emerges in research documentation. Several of the frameworks and perspectives used to probe data as a notion and as something that is a part of human doings and ongoings in research and other areas of work and activity can also be useful for approaching paradata. Like paradata, data is a slippery concept. It is also, and to a much greater extent than paradata, a powerful operationaliser in the scholarly arena as, the ‘substance’ that is managed and curated (e.g., Koesten et al., 2020), described (e.g., Hjørland, 2023), collected, searched for (e.g., Gregory et al., 2019), analysed and interpreted (e.g., Leonelli and Tempini, 2020). Although data is a term often used to denote a general property of scholarly work and a key occurrence in academic research and its ecosystem of related activities, data is also something that can be thought about and approached as a thing (Buckland, 1991). This way of thinking emphasises the physical media, including the material strata that make up digital data (Kirschenbaum, 2008), where the data is recorded or otherwise present. Data as a thing appears in different shapes and across scholarly disciplines and the wider domains of labour and leisure. Beyond the examples of soil samples and data generated in medical examinations offered above, scholarly data can be the recorded interactions of high-energy particles (Knorr-Cetina, 1999), artefacts and observations made and documented during archaeological field excavations (Huvila, 2014), and the reports and publications coming out of already completed studies to be used for data aggregation purposes (Faniel and Yakel, 2017).

The breadth and complexity of the challenges related to discussing what data ‘is’ can be illustrated further. Heritage data is a term used to refer to the holdings of archives, libraries, museums, and other heritage institutions and organisations (Bruseker et al., 2017). Different modes of data work have been observed to be important parts of, for instance, the co-productive activities of videogame players (Sköld, 2017; Steinkuehler and Duncan, 2008; Warmelink, 2013) and people invested in self-tracking their daily, non-labour activities (Abtahi et al., 2020; Trace and Zhang, 2020). Scholarship concerned with defining data has stressed that while it is a complex notion (see e.g., Carlson

and Anderson, 2007), it fundamentally signifies representations ('evidence' as per Birnholtz and Bietz, 2003, no pagination; 'facts' or 'observations' as per Rowley, 2007, p. 170; 'measurements' as per Zimmerman, 2008, p. 633) that describe the characteristics of things, events, and processes (e.g., Beretta, 2024; Rowley, 2007). Although it has been argued that data should always be understood as a social and cultural phenomenon in the sense that no set of data is without a context of production (Birnholtz and Bietz, 2003; Borgman, 2012; Gitelman, 2013; Hjørland, 2018), several attempts have been made to explain what data is by highlighting that data can potentially exist without a conceivable context of use. From this perspective, data can be independent of whether or not conditions are in place for understanding its contents or structure and putting it to use in a relevant way (Rowley, 2007). This is in contrast to a notion like information, which is generally (but not always, see e.g., Wallis et al. 2013) understood to have a higher degree of innate interpretability even though its usefulness and modes of interaction with the setting of use is anything but monolithic and vary widely from context to context (see e.g., Hjørland, 2023).

3.3 Understanding (Para)data in Context

There are numerous studies that further add to the complexity of what data is by stressing the inextricably contextual nature of data. Although this way of thinking about data encompasses an extensive range of perspectives and emphases, data here comes across as something that is always enmeshed with human affairs and that has to be understood by delving not only into the data itself but also taking into account what is done with data, through data, and by data in different settings of activity and work (Borgman, 2012; Hilgartner and Brandt-Rauf, 1994; Latour, 1987). In this view, data becomes data when it is mobilised into the practices that engage with it to achieve some result or to reach some goal (Berg, 1996; Birnholtz and Bietz, 2003; Oliver et al., 2024). These data practices can be those of making and interpreting data (Berg and Bowker, 1997; Huvila et al., 2021a; Huvila et al., 2021c), but also of, for example, adding marginalia (Edwards et al., 2017) and curating, aggregating, or visualising data (Börjesson et al., 2022b; Sköld et al., 2022). Data is not simply a resource in data practices, but an impactful agent that affects the outcomes and procedure of the practice of which it is a part, that is, creating and maintaining boundaries between groups of users and different forms of data use (Brown and Duguid, 1996; Huvila, 2011; Harvey and Chrisman, 1998) including the shaping of ethics and politics (Börjesson et al., 2020;

Olson, 2002) and impacting the ways in which data is organised (Nadim, 2021), circulated (Kansa and Kansa, 2021) and reused (Pasquetto et al., 2019). Some studies underline the sociomaterial nature of data. From this perspective, data's social – relating to the full range of data work and data-related activities – and material – that is, the infrastructural and technological – dimensions have to be considered in tandem when attempting to grasp what data is and what data does in its different settings of use (e.g., Berg, 1996; Pinch and Bijker, 1984). Data and data practices blend, cojoin, and collate into assemblages of social and material data doings that are continuously negotiated and to different extents local in flux (Law, 1993; Pickering, 1995).

Relating specifically to what underpins data identification, the context in which data is produced has been shown to greatly impact what is required for data to be captured and successfully reused by parties other than those involved in its making (Borgman, 2012; Durrant et al., 2011). Data is always entangled with the setting where it came from – a setting which is constituted in turn by certain methodological choices and implementations, theoretical interests, research purposes, institutional factors and technical know-how (Berg and Goorman, 1999; Faniel and Zimmerman, 2011; Van House, 2002b). If data is thought about and put to work in an environment framed by other epistemic horizons and approaches, challenges may arise that impact the usability of the data unless mitigated. Examples of challenges include difficulties in understanding what the data signifies, how it can be interpreted, and to what extent it is purposeful and trustworthy (Baker and Yarmey, 2009; Faniel and Jacobsen, 2010; Rolland and Lee, 2013; Yakel et al., 2013). Using various kinds of data descriptions in order to elucidate what is considered to be, for identification and usability, the central elements of the data is considered to be one of the main ways to bring the horizons of data makers and data reusers closer together. Thus, data useful in one setting can also be useful (albeit possibly in different ways) in another (Baker and Yarmey, 2009; Carlson and Anderson, 2007; Zimmerman, 2008). Data descriptions that decrease the 'distance-from-origin' (Baker and Yarmey, 2009, p. 13) between data makers and data reusers can be free-form or structured to varying degrees using, for example, standards and recommendations. However, all have to navigate variations on the same dilemma: From one data reuse scenario to the next, it is difficult to know what information is required to bridge the gap between the locales where data is first made and used and the secondary-use contexts (Carlson and Anderson, 2007; Faniel and Zimmerman, 2011; Zimmerman, 2008). Additionally, metadata schemas and standards that are deemed to be purposeful in one domain may not work equally well in another (Birnholtz and Bietz, 2003; Fear and Donaldson, 2012; Yakel et al., 2013).

3.4 Tracing the Interplay between Research and Research Documentation

For paradata created and generated during the course of research work, research documentation is one of the main sources from which it can be harvested. Identifying documentation as a source of paradata reflects a fundamental tenet of sociomaterial insight into the practices and processes of scholarly knowledge-making: namely that documentation is something present and notably, but differently (Becher, 1989; Knorr-Cetina, 1999), involved in all major steps of the research lifecycle (see e.g., Latour and Woolgar, 1979; Trace, 2011). The overall centrality of documentation and systems of documentation in human epistemic endeavours is also highlighted by research that examines how knowledge is made and negotiated in broader contexts of labour and leisure beyond the scope of science (Harper, 1998; Law and Lynch, 1988; Levy, 2001).

It has been similarly noted that research documentation is valuable for purposes of paradata collection and use (Geiger and Ribes, 2011; Hodges, 2021; Sköld, 2018). Documents and documentation can manifest in many different forms and contain information about any number of things, as is shown by a range of information studies research (see e.g., Briet, 2006; Lund, 2009). The ubiquity and importance of documentation in knowledge-making stems not principally from document 'aboutness' (Brown and Duguid, 1996; Hjørland, 2000), although aboutness has been underlined as being of importance (e.g., Buckland, 2018), but rather from how documentation functions epistemically and socially (Lund, 2010).

In academic research settings, documentation can vary in how it functions and appears across disciplines. It is understood to be a material, historically situated, and genre-bound resource in scholarly work that simultaneously and in various ways supports, coordinates and perpetuates many of the defining practices of research (research design, data collection and analysis, ethical constraints, the reporting of findings and results; see e.g., Cragin and Shankar, 2006). Documentation is also the principal scholarly deliverable of science in the form of books, papers in scholarly journals, and datasets (Frohmann, 2004a; Knorr-Cetina, 1999; Trace, 2011). In addition to emerging as scholarly end products, the major part of research documentation is composed of documents created for supporting purposes during the course of scientific activity. Examples include field diaries, notes and sketches, and other scholarly by-products and marginalia (Edwards et al., 2017; Huvila et al., 2021a; Shankar, 2009; Spedding and Tankard, 2021).

Research examining the production of scholarly documentation during scientific work has identified overarching issues affecting the extent to which

documentation offers useful and usable paradata. One such issue is that a large part of the documentation arising from research activities is created to serve project- or research-specific purposes, which often do not consider the need to support secondary data use by any party not already involved in the original data production (Börjesson et al., 2022b; Sköld et al., 2022). For paradata present in research documentation, this means that access and identification of supporting resources to be able to interpret the paradata in sufficient context are likely to be pervasive challenges.

Another issue concerns the ability of expert insiders to explain data work and appearance. A major obstacle for identifying and using paradata in research documentation is the bridging of ‘insider’ and ‘outsider’ horizons of data insights and understandings. Sometimes, even expert insiders have trouble explaining what precisely their work with data looks like and why they do the things they do (Ciborra and Lanzara, 1994; Shepherd and Rudd, 2014). This is especially the case if the audience belongs to disciplinary domains or subdomains with different epistemic and methodological hallmarks (Faniel and Yakel, 2017; Niu, 2009; Pardo et al., 2006) or levels of in-field proficiency (Faniel et al., 2012; Yakel et al., 2013). The degrees of similarity and difference between the contexts of paradata making and paradata use are important broad-scope parameters in explaining the potential and challenges of paradata localisation and use (Borgman, 2012; Faniel and Zimmerman, 2011). However, studies of facilitating factors and barriers of data reuse among, for example, archaeologists (Faniel et al., 2013), ecologists (Zimmerman, 2008), earthquake engineers (Faniel and Jacobsen, 2010), medical researchers and space physicists (Birnholtz and Bietz, 2003) have shown which paradata characteristics researchers find useful for facilitating understanding of what a dataset manifests and how it can be engaged with secondary-use purposes. Paradata containing information about how a dataset has been collected, processed (Borgman, 2012; Börjesson et al., 2022a; Rolland and Lee, 2013), and organised (Börjesson et al., 2022a; Yoon, 2014a) – including explanations of variables (Rolland and Lee, 2013) and concepts (Faniel et al., 2012) with a high degree of explanatory potential – emerges as valuable in several settings of use, although there are variations in the findings across themes and emphases.

Similarly useful is paradata underpinning facets such as the background and goals of the research producing the data (Borgman, 2012; Niu, 2009). Alongside paradata quality, coverage, findability and completeness impacting the success of data reuse attempts (Faniel et al., 2016; Faniel and Yakel, 2017) are the personal elements of the individual data reusers. Examples of personal elements include the degree of required data-related skills (Niu, 2009) and literacies (Börjesson, 2021; Kim and Yoon, 2017), and perceptions of data usefulness,

sufficiency of resources supporting the data reuse venture, degrees of trust in the data (Faniel and Jacobsen, 2010; Yoon, 2014a), and the credibility of the data and its makers (Faniel et al., 2016; Franks, 2024; Huvila, 2020).

Several studies have adopted different approaches to determining how paradata can be identified in research documentation of both the end product and supporting resources. These studies show that paradata can be identified in the full range of digital and physical research documentation, from papers to reports and monographs, from notes in databases and datasets to diaries and codebooks (e.g., Börjesson et al., 2022b; Niu, 2009; Sköld et al., 2022), as well as the visual representations (including photographs but also other types of visual representations like maps and drawings) underlined as important representations of scholarly practices and processes by Huggett (2023) and Huvila et al. (2023a). The breadth of documented paradata corresponds to how researchers document research practices and processes, something which is done in many ways, for many purposes, using many different means (Huvila et al., 2021c; Huvila et al., 2022a; Knorr-Cetina, 1999).

The paradata itself can also emerge differently. Börjesson et al. (2022b) note that paradata can be either explicit and manifested in descriptions of how certain operations impacting the resulting research data were planned and executed, or indirect evidence by which data actions, deliberations and processes can be traced (see also Huvila et al., 2023a). Due to the potential richness of paradata stemming from the many traces of data practices and processes in datasets, it has been suggested that 'messy' and non-harmonised datasets might not be a hindrance to data reuse (e.g., Richards et al., 2021) but actually an asset in facilitating insight into what measures were taken in its creation and preparation (Börjesson et al., 2022b). Paradata can, for instance, be identified in missing information and incompleteness that, while often being detrimental to paradata quality (West and Sinibaldi, 2013), via information gaps and otherwise non-available and non-interpretable information, can be used to some extent as useful paradata due to it reflecting the priorities, methods, and abilities supporting the original data-making endeavour (Huvila et al., 2023b; Ullah, 2015). Normalised and harmonised datasets are easier to make interoperable, which would open up opportunities for identifying and aggregating paradata in large corpuses of research documentation (Gunnarsson, 2020; Kintigh, 2006). There is a risk of over-standardising datasets (Plantin and Thomer, 2023), however, when in different ways adapting them to standards for enhanced interoperability (Maron and Feinberg, 2018). In this way, dataset normalisation, harmonisation and standardisation can mean that potentially useful paradata is cleaned away in the data curation process (Börjesson et al., 2022).

3.5 Following Practice: How Archaeologists and Archaeological Research Data Professionals Identify Paradata

The interview study of how archaeologists and archaeological research data professionals use and create paradata exemplifies how information qualifying as paradata provides information about a varied set of scholarly practices and processes, and is heterogeneous in the sense that it emerges in research documentation with different characteristics and access affordances. The following section will examine the three types of documentation that were most commonly consulted by the interviewees as sources of paradata, and illustrate the descriptive qualities and what practices and processes that the paradata that was obtained from these types provide information about. The document types are: method descriptions, references and structured information. They occur in several interconnected forms and shapes across many types of research documentation, as shown in Table 3.1.

3.5.1 Method Descriptions

The interviewed researchers and data professionals commonly consult various types of ‘method descriptions’ in the documentation when working to identify paradata in research documentation. Method descriptions occur in several kinds of research documentation and are differently manifested in terms of how extensive the descriptions are and where they can be located and accessed.

Table 3.1 *A selection of the documentation types that the researchers and archaeological research data professionals in the CAPTURE interview study identified as paradata sources*

Documentation type	Documentation type description	Prevalent instances of research documentation
Method descriptions	Accounts of the actions and deliberations involved in creating research data	Journal articles, monographs, reports, notes, working logs, email messages
References	Codes or other identifiers that serve as connectors between items of research documentation	References to scholarly publications, datasets, authors, data authors
Structured information	Information organised according to a standard or other organisational principles	Standards, local standards, data dictionaries, domain ontologies, scripts and code

Method Descriptions in Reporting Documentation

And I just thought, "Oh no, [the georeference coordinates are] wrong".
[- - -] But then I went and found the original paper and I checked
the coordinates and the coordinates in the database are actually the
coordinates that are in the paper, they were just wrong. But the paper
was from 1992. So, people were not using GPS then. (Interview no. 7)

Researchers and data professionals ubiquitously use method descriptions in published (journal articles, monographs) and unpublished (reports, manuals and guides, theses and dissertations) scholarly outputs as sources of paradata. Method descriptions in these types of scholarly reporting documentation are often easy to identify and examine, with the exceptions of older and undigitised items. They are fairly homogenous in the sense that descriptions account for the main elements of research processes by reflecting the guiding research questions and framing of the inquiry or task, methods and materials, interpretative resources, and results in method and materials sections and similar passages. The paradata identified in the method descriptions of scholarly reporting documentation is broad in scope and reflects the practices and processes of how the data has been created, used and curated in a way that is narratively and semantically comprehensible. To some extent, it is also formalised and genre-bound, although the scope, detail and format of the paradata can vary. Some reporting documentation is highly standardised, while other method descriptions are free-form and provisional, reporting observations and reflections that stem from different forms of investigative and analytical activities, including fieldwork. Notes and field notes of this nature are not the final reporting documentation of the research processes they come from, but are an important building block for it. They contain descriptions of how research methods have been employed in empirical research and what results emerged out of the process.

Paradata about the procedures and means involved in creating primary (previously non-existing) data in an empirical setting like a field site or laboratory, including instruments and instrument settings, are often identified in the method descriptions of reporting documentation. Data-use paradata is also present there and offers insights into how the data has been analysed to support the results presented in, for instance, the article or report. The main elements of management paradata available in the method descriptions of reporting documentation provide information about the procedures involved in the aggregation and integration of available data resources. Examples of such sources include bodies of source documents and collections of already created research data that are aggregated into composite datasets and databases.

Method Descriptions in Auxiliary Documentation

And those sorts of things, decisions and exceptions and strange stuff, that is what we document while talking about it, and by having that discussion be part of the [project] records. (interview no. 19)

Paradata can also be identified in the methods descriptions of auxiliary documentation. While paradata in the method descriptions of reporting documentation emerges from the relationship between the documentation and the modes of scholarly work reported there, the researchers and data professionals also identify paradata in method descriptions of other types of documents. Rather than being one of the final outputs of a research process, these types of documents have intermediate or supporting characteristics. They are created and used to document, guide and otherwise support the research being done. Auxiliary documentation of this kind encompasses a broad range of research documentation, from highly structured and detailed files to sparse hand-written notes and documentary fragments.

The method descriptions where paradata can be identified in auxiliary research documentation similarly vary in comprehensiveness, format and accessibility. Comprehensive method descriptions in auxiliary documentation are often similar to those present in reporting documentation in that the descriptions have narrative structures and are written for audiences other than the author. In contrast to reporting documentation, however, method descriptions in auxiliary documentation are usually primarily intended for internal or project-specific applications. As a result, they are less formalised and have, for instance, a substantial presence of abbreviated forms of notes and comments, and they generally require more domain knowledge to understand and use.

A recurrent type of auxiliary documentation with comprehensive method descriptions is task and procedure documentation. This documentation can be shared working logs of project team members, in which paradata is present in the form of research activities formulated and tracked objectives, documentation of data management procedures and underpinning decisions stored in joint online resources such as internal wikis or repositories.

Method descriptions present in the entire body of notes and drafts produced during the course of scholarly work can also be used to identify paradata. These method descriptions come in the form of sketches that outline implemented methodologies or workflows, records of how a dataset has been analysed, enriched with metadata or otherwise managed. Email messages are another source of comprehensive method information consulted to identify paradata and, likely due to the prevalence of email communication in scholarly

work, paradata about a broad range of research activities can be identified there – from paradata about how research data has been categorised and described to the steps of data analysis and creation.

Less comprehensive method descriptions in auxiliary documentation are characterised by often being produced only to support the person writing them. They are very informal and use large amounts of shorthand and abbreviations. The method paradata in less comprehensive auxiliary documentation is not narrative but trace-like, representing isolated actions or disassociated pieces of method paradata, and it requires a high degree of domain familiarity to be informative. Examples from the interview study include communicating the lack of precision in the geolocation of finds by putting three zeroes at the end of the related GIS coordinates, and using certain signifiers like dashes and question marks to indicate interpretative uncertainty in dataset items.

3.5.2 References

References in Scholarly Publications

And if it's something quite specific and well-defined, then those citations tend to be really solid and [...] like, okay, you know, here is the canonical article that establishes why this particular term is something that we record [in the dataset]. And then some things that are like, okay, well, that's just kind of general domain knowledge, those tend to get citations in kind of introductory handbooks to, you know, the handbook to archaeological specimens or whatever it is. (interview no. 12)

Another commonly consulted source of paradata are diverse formal and informal references. References in research documentation function in interlocking ways: either by the reference itself offering insight into research practices and processes, or by referring the researcher to the publication or referenced item where the paradata can be identified.

Paradata emerging from references in the first instance is only accessible with significant domain familiarity. A reference to landmark publications with large impact on subsequent works produced in its line of research can signal that the research output where the reference is present conforms to certain methodological traditions or modes of thought. Additionally, a reference to well-known datasets in a specific area of research – one example from the interview study is *Kong Valdemars Jordebog*, a Danish fourteenth-century census book containing, among other things, land ownership information not available in other sources – might also provide trace information about how the

data has been created, curated and used for study and what contextual elements have impacted the dataset.

In the second instance, researchers identify paradata in references by engaging with the items that the references point at. These items can be published or unpublished texts, datasets or parts of a datasets and not seldom they provide paradata of a narrative nature, such as method descriptions. Access affordances and the paradata provided vary between different kinds of references, and references present in datasets and in scholarly literature, respectively, are the most prevalent kinds.

References in Datasets

Well, there are find coordinates [in the aggregated dataset] and included are also three additional types of references to other data sources: the ID of find post in the [Swedish National Heritage Board's database for archaeological sites and monuments] if it's available [- - -], a reference to publications appended to the data [. . .], and a museum collection inventory number – that's the most important thing, so that the [physical finds] can be located. (Interview no. 27)

The references present in datasets that researchers and data managers use to identify paradata refer to both literature and other datasets. References to other datasets often occur in datasets that themselves are the result of aggregation or integration data work, where multiple data sources have been merged or otherwise connected into a structure that can be consulted as a whole. These references are used by researchers to identify paradata about the individual datasets to better understand the research work that created them, or to understand how the aggregated dataset has been assembled. There are also internal references that link the segments of a dataset, and which can offer insights into how the dataset and its parts are related to each other. Datasets in database environments have tables that are linked together by identifiers, and beyond offering search functionalities these links can be used as paradata because they provide information about how the data in the tables is connected. Examples include how the results presented in one table connect to underpinning method descriptions or other research process elements in another table.

References in datasets to literature are employed to identify similar paradata, namely information about how the dataset has been created including definitions of variables and descriptions of stances and decisions taken. In addition to this, references in literature are also used to learn how related datasets have been created, used and curated. Most commonly, this literature reports the

results of research based on the data in question, as is explained in the previous section. References embedded in the method descriptions can offer more detailed insight into the scholarly processes that shaped the supporting dataset. References in scholarly literature that refer to datasets can be used in order to gain access to the paradata in the dataset itself, including gaining a better understanding of how the data has been described and engaged in the analytic process.

References to Authors and Data Authors

Yeah, I mean I would say probably the most important information that I would add to this project [dataset] description is the credits [of the data author]. (interview no. 18)

A common position among the researchers and data professionals interviewed in the study is that references to the actors involved in data authorship – names as well as laboratory affiliation of the individuals – are highly useful to elicit paradata. Although the name alone of a data author can provide information about the contents of a dataset and modes of work involved in producing it, the main way to gain paradata through data authorship references is to communicate with data authors. Data authors can give access to several important paradata sources, such as information about the techniques, methods, software and instruments employed, as well as other rich context and process information. Advantages to sourcing paradata via authorship references are that the data authors can explain facets of the research process that are rarely documented and otherwise difficult to obtain. Such facets include the exploratory parts of the research process that are seldom reflected in the final research outputs, and underlying disciplinary norms and domain knowledge that are difficult to discern and articulate but with impact on the data created.

3.5.3 Structured Information

'Structured information' is the third documentation type of great importance for the modes of paradata use found in the interview study. The 'reference' documentation type is fairly homogenous by being mainly of a trace rather than a narrative character, and being principally formalised. The structured information used to identify paradata, on the other hand, is akin to the 'method descriptions' documentation type in being considerably varied. Structured information that can be used to access paradata range from technical and descriptive standards to different degrees formalised local conventions for

describing and presenting data including scripts and code, and semantic typologies like data dictionaries and domain ontologies.

Standards and ‘Local Standards’

I know that for [the] CIDOC CRM [standard], there are many options with that ontology to document procedural information such as who did what, when, with what data. (interview no. 11)

Researchers and data professionals identify paradata in research documentation by looking for standards employed and referenced. Standards are used as interpretative tools that, due to their nominal status as institutionally backed and highly organised descriptors of how scholarly actions should be performed – CIDOC CRM and geolocation standards are the most commonly occurring examples in the study – can be used to gain insight into settings of scholarly data work across sites, disciplines and time. Standards can yield paradata about how datasets have been described by providing the supporting metadata principles. They can also show how the components of an aggregated dataset have been linked together through standards that establish and describe research tasks, processes or other categories present in the integrated datasets. Standards are also consulted in order to identify paradata about how data has been analysed and created, including how instruments and computational tools have been calibrated and what their operating conditions were like during the time of data creation.

Apart from standards, there are many other established ways of creating, using and managing data that are described in research documentation, and are sought by researchers to identify paradata. These ‘local standards’ vary more in the extent to which they are formalised and documented than do institutionalised standards. They also have more local application: in complex and large-scale research tasks as well as more limited ones, local standards are sometimes used in the data collection and description work of individual researchers. Occasionally, they coordinate the scholarly efforts of researcher groups or networks. The paradata provided by local standards is similar in type to the paradata that can be identified in standards, but showcases a broader scope in that, in addition to data on data creation, description and management, research contexts and modes of data use, it provides a greater extent of narrative and trace paradata about group attitudes, idiosyncratic approaches to different scholarly tasks. Local standards have similarly diverse access affordances that can be described sparingly or comprehensively, and can be found in the often internally kept auxiliary documentation produced within research projects and accessible in online repositories and other online resources.

Semantic Typologies

Yes. I am quite specific on that, because I think it's very important to be specific on what kind of terms and stuff you use, also so other people know exactly what you mean when you call something "orange". Or when you call something "angular", what you mean with it. (interview no. 8)

Semantic typologies of different varieties are another prevalent form of structured information in research documentation used to identify paradata. These typologies offer paradata that is, on a general level, similar to that identified in standards and local standards – it tells the reader about the main phenomena that exist within the area that the typologies encompass and details what the relationships between the phenomena are in a way that is often largely formalised. In contrast to standards and local standards, the semantic typologies are principally oriented towards mapping areas and domains of conceptual knowledge and are less concerned with the practices and processes of how research data is created, curated or used. Semantic typologies are useful sources of paradata, providing insights into the scope, pivotal concepts and relationships between concepts that have informed data work in research projects and research data collaborations. This paradata presents the opportunities to better grasp the epistemic horizons of a certain research venture, and enables purposeful comparisons and data integration, by providing definitions of notions and parameters that signify one significance of a range of possible significances.

Semantic typologies in research documentation that have been collected or managed over an extended period of time can also offer paradata about research procedures, where changing ways of describing data or defining terms and parameters traces reconsiderations and other changing modes of work. Instances of semantic typologies found in research documentation are data dictionaries, which provide definitions of impactful terms, variables and parameters present in a dataset; domain ontologies, which frequently map high-level entities comprising a certain data domain; and scope notes, which offer paradata about the area of empirical reality that the dataset describes along with the main conceptual vocabulary and methodologies used to interpret and study it.

Codes and Scripts

So, I expect the reader to start in the R Markdown file and then kind of navigate through the rest of the compendium according to the code that I've written in there. And perhaps, like reverse the past, reverse engineer my analysis from the R Markdown, which is kind of the recipe that brings all the bits and pieces together. (interview no. 30)

Codes and scripts are an additional instance of structured information used to identify paradata in research documentation. These are written, for example, to analyse, visualise, aggregate or otherwise process and use research data using, for example, Excel formulas, and code in the R or Python programming languages. The paradata in codes and scripts shares characteristics with paradata in standards, local standards and semantic typologies. Even though the code or scripts themselves may differ in terms of how well documented they are and to what degree commands and functions have been stringently implemented, the programming and scripting languages are in a fundamental sense documented and described in a way that is meant to transcend different areas of operation and implementation.

Code and scripts – especially codes and scripts that have been written to be transparent – are also consulted to trace what operations were performed on the datasets involved, and to understand why the resulting outputs look the way they do. These operations can provide information about established modes of work in a manner similar to standards and local standards. They may also make it possible to determine what parameters or variables were defined in the preprocessing stages of scholarly data work and to see how they support in the resulting analysis or visualisation.

3.6 How Can Paradata in Research Documentation Be Identified, and What Does It Look Like?

The basic premise of this chapter is that knowing how to identify data useful for solving the task at hand, and being able to know what this data might look like, are important conditions of successful (para)data procurement. The observations from the CAPTURE interview study add to the significance of this starting point by reinforcing the characterisation suggested by previous research that the work of science to large extents resembles work of and with documentation (Frohmann, 2004a; Shankar, 2009). The interview observations also show that paradata of many kinds can be identified in a broad range of the scholarly documents and snippets of documentation like out-of-context notes and potentially informative traces of acts and events available in such documentation (see also, e.g., Huvila et al., 2021a; Huvila et al., 2023a; Niu, 2009).

From this outset we can examine how paradata in research documentation might be identified, and to describe paradata's potential appearance from the two perspectives of paradata as thing and paradata as practice. The former perspective directs attention towards how paradata is physically manifested in

research documentation; the latter emphasises how paradata is involved in and emerges from the modes of work and activities enacted by the interviewed researchers and data professionals. Together these perspectives highlight related but distinct facets of paradata available 'in the wild' of research documentation, how it might be identified, and what it might look like.

3.6.1 Paradata as Thing

When approaching the instances of paradata described in the interview study from the data as thing perspective, paradata emerges as method descriptions (in e.g., journal articles, monographs, unpublished reports and dissertations, and in notes, work logs and emails), references (in and between e.g., datasets and publications), and structured information (in e.g., standards, local standards, code and scripts) examined for their information about past research practices and processes.

Table 3.2 summarises where paradata is identified in research documentation and provides examples of key sources to consult in the capture of paradata. The table also describes the overarching characteristics of the paradata that can be extracted from these sources alongside the sources' access affordances. In the table, 'paradata access affordance A' signifies that the paradata sources are accessible in online public access catalogues, publication or preprint repositories, or open data repositories. 'Paradata access affordance B', on the other hand, shows that the paradata sources are accessible by consulting data authors or data authoring organisations. If the paradata sources require significant domain knowledge to access they are marked as having 'paradata access affordance C'. Access affordances referenced within parenthesis signifies that it is present to a lesser extent.

Corresponding and complementary to the results of earlier inquiries into paradata in research documentation (Börjesson et al., 2022b; Huvila et al., 2021c; Huvila et al., 2022b), the document types emerging from the analysis in the preceding section contains several kinds of paradata. The kinds of paradata range from paradata about how research data was created, including both machine use and research steps and methods employed, data aggregation and curation and analysis procedures, to how the data has been organised and structured. Although method descriptions, references and structured information alongside the prevalent instances of paradata manifestations they collate are examples of paradata identifiers signalling that the presence of paradata can be reasonably expected, the paradata may vary considerably in terms of aboutness. That is to say, there might be data creation paradata, data curation paradata, data organisation paradata and so on but the paradata may also

Table 3.2 *An overview of how to identify and access paradata in research documentation based on the results of the CAPTURE interview study*

Where paradata is identified	Key sources to consult	Paradata characteristics	Paradata access affordances
In method descriptions	Published reporting documentation like journal articles and scholarly monographs	The paradata describes the principal activities and resources involved in data creation, curation and use	A
	Unpublished reporting documentation like reports, manuals, guides, theses and dissertations	The paradata is akin to the paradata in method descriptions of published reporting documentation, but more varied in comprehensiveness and content	A, B
	Auxiliary research documentation like notes, work logs, drafts, email messages, internal documentation of research procedures	The paradata describes often discrete activities or resources involved in data creation, curation, and use with significant variances in comprehensiveness, structure and format	B, C
In references	References in published and unpublished reporting documentation; the referenced literature	The paradata describes a wide range of paradata principally informing about what literature is discussed with and drawn upon, and in what respects	A
	References in datasets including databases; the referenced datasets and databases	The paradata describes how datasets are linked to each other and to literature, and how dataset components are linked	A, B, C
In consultations with data authors	Data authors and data authoring organisations	Data authors or data authoring organisations can provide broad-range paradata that partly may not be otherwise accessible	A, B
In structured information	Standards in published and unpublished reporting documentation; the referenced standard	The paradata describes how data has been described, structured, created, curated and used	A, B

Table 3.2 (*cont.*)

Where paradata is identified	Key sources to consult	Paradata characteristics	Paradata access affordances
	‘Local’ standards in published and unpublished reporting documentation	The paradata describes how data has been described, structured, created, curated and used, but will significant variances in comprehensiveness	A, (B), C
	Semantic typologies like ontologies, data dictionaries, scope notes, concept mappings	The paradata describes the significance and scope of concepts and conceptual structures	A, (B)
	Codes and scripts like R code, Python code, Excel formulas	The paradata describes detailed data creation and processing steps	A, (B)

exhibit a range of other characteristics along intersecting gradients of formalisation, comprehensiveness and scope. The paradata can be highly formalised as shown, for instance, in data created or described using geolocation or other standards, or be free-form in structure and content as in notes and sketches intended for use within a research project or group. Both standardised and free-form paradata is rule-bound, however, albeit by different sets of constraints and with different degrees of formalisation; ‘informal’ documentation like email messages, notes and notations in databases may also be tied to paradata genres and adhere to localised and more idiosyncratic standards or ways of organising, relating and documenting terms, variables or research procedures (Börjesson et al., 2022b; Huvila et al., 2022a).

As observed by Maron and Feinberg (2018) and Birnholtz and Bietz (2003), data impacted by standards and guidelines may also operationalise these in different ways and to different extents, resulting in paradata that may be less formalised than expected. Similarly, paradata in certain document types and in certain regards appear more comprehensive than others, such as method descriptions in reporting documentation vis-à-vis trace-like notations of actions in a work log. It may be that while the method descriptions better adhere to the established way of reporting how a study was planned and enacted, the work log may yield less processed and, within a more limited scope, fuller and better serialised expressions of what was done during the course of a study.

3.6.2 Paradata as Practice

Paradata in research documentation can also be considered from the perspective of practices. In the interviews there are two practices that particularly come into play: how archaeologists and archaeological data professionals create paradata in research documentation, and how they identify it.

When it comes to the practice of creating paradata, several things can be observed. Firstly, paradata can be created with the objectives of several audiences in mind (Faniel and Zimmerman, 2011; Huvila and Sköld, 2023). While it may be difficult to clearly map categories of underpinning intentionality to what paradata emerges from different document types, there are some patterns that are striking. Paradata in the method descriptions and references of published reporting documentation (monographs, journal articles) are more likely to be created with an audience in mind that is external to the local setting of the reported research task. The narratives of what activities, decisions and deliberations took place, and how as well as when, provided by this paradata can be expected to be, to some extent, shaped by the regimes of scholarly discourse and expectation that vary across research traditions and disciplines (Huvila et al., 2021c; Shankar, 2007).

There is also paradata that is similarly embedded in the context where it was created, but where the hallmarks of its embeddedness are distinct – and akin. This paradata has strong processual characteristics in that it was created to support a particular process, for example, data analysis, aggregation or organisation. It can be identified in material such as notes and shorthand annotations in datasets, and in data structures that correspond to modes of work that are exercised within very local circumstances where the intended paradata stakeholders are few, comprising only the paradata creator and possibly members of a research group.

It can also be observed that paradata creators document paradata using a range of strategies. These strategies involve different document locations, approaches, and means (Huvila et al., 2022a; Sköld et al., 2022). Some strategies produce results, such as references and method descriptions in published outputs, data work informed by standards, or more regimented unpublished items like reports, that vary less in terms of what the paradata looks like and how it can be identified. Others display a larger degree of idiosyncrasy and variety.

In the interview study, paradata that describes sequences of events, even if described rudimentarily or with little narrative structure, is more often self-contained and available in documentation separate from datasets or publications. Paradata documented in formats more closely resembling traces and

marginalia tends to be present in a broader range of locations in and across datasets and collections of documents (Edwards et al., 2017; cf. Sköld, 2017). In many cases, there is some degree of structure to paradata traces and marginalia, which can be used to identify and interpret them. This structure is often anchored in local modes of documenting research practices and processes that are themselves to a lesser degree documented or institutionalised beyond their immediate context of creation.

The interview study also highlights some patterns in the practices of identifying paradata: the interviewees identify paradata among data authors and contributors, in the gaps and relationships between research documentation, and in research documentation. Consulting data authors and contributors is a crucial strategy in the practice of identifying paradata. It is not always possible to carry out, for example, because data authors or contributors cannot be contacted or because it is difficult to determine, especially in large datasets, who was responsible for what data task or operation. While consulting the people involved in data creation, curation and use is an approach that has the potential to yield detailed and comprehensive paradata, it is not always successful even if contact between paradata maker and paradata user can be established. The interviews reveal that the questions meant to facilitate locating and using paradata risk being framed within the horizon of paradata use to the extent that they are difficult to answer from a data creator's perspective. The reason for this is that questions can be closely tied to, for example, the research objectives, methodologies and modes of reasoning pertinent to the inquiry that the paradata is supposed to support, and these can be very different from the research setting from which the paradata originated (Rolland and Lee, 2013; Van House, 2002b; Voss, 2012).

Paradata is also identified in the gaps and relationships of research documentation, in the sense that the relationship between different parts of the same resource, for example, multiple entries in a database, or between several kinds of documentation can outline past data events. This practice is very useful in that it can be used to grasp undocumented or not insufficiently documented elements of data work in a broad range of documentation types. This can be done regardless of their degree of formalisation, comprehensiveness and scope, by serialising and tracing change and constants in all of the documentation types discussed above and presented in Table 3.2.

Another way of identifying paradata 'in between' research documentation resembles the 'reputational cues' that Huvila argues inform credibility assessments of research documentation (2020, no pagination; see also Berg and Goorman, 1999; Faniel et al., 2016; Fear and Donaldson, 2012). This method involves considering if there are some elements (the data author, the research

project the documentation comes from, the documentation itself) that are institutionalised or otherwise well enough established within their domains to allow for making inferences regarding what the data is about, and how it has been accumulated, created or managed. To determine if such inferences can be made, it is necessary to have sufficient domain insight and understanding of the social and practical circumstances of the research work in order to be able to determine if there are reputational cues and what paradata they might offer.

In the same way that it is a common strategy to consult paradata in different documentation types that address relevant but not identical things (Börjesson et al., 2022a; Niu, 2009; Yoon, 2014b), it is rare to use discrete approaches to identifying paradata. Instead, paradata is often identified by consulting data documentation and data authors and contributors in some combination, and attempting to discover patterns in the research documentation that can offer useful insights into the scholarly data processes of interest (Fear and Donaldson, 2012; Huvila et al., 2021c). The interview study similarly shows that paradata from one source is rarely sufficient to answer all process queries. The usefulness of paradata rests to a significant extent on the degree to which it can be connected to other paradata or to what is known about how it is created and commonly used in the field of interest.

3.6.3 Paradata as Thing, Paradata as Practice

The two perspectives on paradata explored above are simultaneously distinct and connected. As a thing, paradata emerges in different forms and with varying scope, comprehensiveness and degrees of formalisation across many of the document types produced during the life cycles of research enterprises – from preparation to reporting, data management and sharing. It is useful to approach paradata as ‘a thing’ for the purposes of identification and localisation, but it is also useful for the same purposes to consider paradata as being a connected data entity that – in line with numerous studies in data and document scholarship (Borgman, 2012; Frohmann, 2004a; Latour, 1987; Shankar, 2009) – is integrally a part of practices of paradata creation and use.

While Chapter 7 continues the theorisation of paradata as a thing-and-as-practice by tying in to the thoroughgoing notion of working knowledge, encompassing both practices and paradata phenomena, it can here be concluded that having insight into the regimes of scholarly discourse and work that impacts how paradata is created and manifested in scholarly documentation is likely to be helpful in knowing how paradata can be identified and what it looks like in a certain discipline or area of research. Together with familiarity of how and where researchers in the domain of interest locate paradata,

knowledge of paradata practices emerges as an important resource in the interpretative work of understanding and using the identified paradata as documentation of past research events and processes.

The discussion of paradata as 'thing' or 'practice' also ties in to the distinction made in Chapter 6, exploring the issue of how paradata can be managed in data repositories, between 'core' and 'potential' paradata. Core paradata refers to paradata that is generally understood as such, see, for example, the 'method descriptions' data type outline above or the more well-defined use of the paradata concept emerging in survey research as shown in Chapter 2. Potential paradata, by contrast, is information about research practices and processes that is not purposely created as paradata, but that might be used as paradata in particular circumstances. Examples drawn from the present chapter include auxiliary documentation, which potentially can be consulted for paradata but that is made for the purposes of supporting research processes. The thing-practice framework used to discuss paradata in this chapter is principally a theoretical tool that enables paradata to be discussed as a data entity that is identifiable from a nominally domain-agnostic vantage point, while retaining the non-essentialist stance that paradata is something that is ultimately sought, identified and put to use within the auspices of particular practices. The distinction between core and potential paradata signifies the extent to which certain paradata categories are conventionalised. The two frameworks, however, inform each other. When paradata is discussed in a thing-y way, it is likely also 'core' paradata in the sense that it is mobilised as such in a broad range of research and data management practices. The paradata potential of 'potential' paradata, on the other hand, is also ultimately based on the extent to which it is – or has the ability to become – a part of the activities and processes taking place within either data management or research-aligned practice.

The framing of paradata as 'thing' *and* as 'practice' also directs attention towards technical and epistemic usefulness thresholds that are relevant for identifying and using paradata in research documentation (cf. the access affordances presented in Table 3.2). The thresholds are connected to each other but emphasise different fundamental conditions of successfully consulting research documentation to identify and access paradata to learn about past data work. This learning is done by navigating the network of 'participants, practices, artefacts and social arrangements' that make up the social and technical underpinnings of scholarly knowing (Van House, 2002a, p. 111; see also Kim and Yoon, 2017 and Huvila et al., 2023a).

Passing the technical and epistemic usefulness thresholds would mean that the paradata identified and harvested can be used to serve the purposes for which it was sought out, which will vary from one data reuse scenario to the

next (cf. Pasquetto et al., 2019). The technical usefulness threshold represents baseline possibilities of accessing and interacting with paradata in research documentation (see e.g., Börjesson et al., 2022b; Niu, 2009; Wallis et al., 2013). It involves having the appropriate software tools for opening and browsing the research documentation, including the means to circumvent issues relating to proprietary or legacy data formats, and the overarching issue of gaining access to the documentation in cases where it is not openly or otherwise available. Given the wide distribution of research documentation across media and locations shown in the CAPTURE interview study and in previous research (Börjesson et al., 2022b; Faniel and Yakel, 2017; Sköld et al., 2022), the access affordances can vary between documentation types.

The epistemic usefulness threshold, on the other hand, underlines the degree of affinity between the epistemic horizons of paradata creation and paradata use. In organisational terms, the former is often related to the original project or venture where the data and paradata were created, and the latter relates to data reuse attempts at a later point in time, during which researchers are seeking to understand how the data came into being, and how it was curated and used. While matters of documentation access and interaction expressed in relation to the technical usefulness threshold are complex and span a range of issues and conditions depending on, for instance, who tries to identify what research-documentation paradata for what purposes, the epistemic usefulness threshold ties into the mechanisms of knowing as they relate to understanding, which arguably has a different kind of complexity attached to it. Familiarity with the context in which the paradata was created (discipline, mode of research, theoretical auspices; Baker and Yarmey, 2009; Berg and Goorman, 1999; Faniel and Zimmerman, 2011; Sköld et al., 2022), methodologies and techniques employed (Faniel et al., 2013; Fear and Donaldson, 2012; Huvila and Sköld, 2023), basic dataset descriptors (scope and provenance Borgman, 2012; Börjesson et al., 2022a; Rolland and Lee, 2013) how the data has been organised (Birnholtz and Bietz, 2003; Börjesson et al., 2020; Maron and Feinberg, 2018), and with the data-creating organisation, researcher or research team (Faniel and Jacobsen, 2010; Huvila, 2020; Yakel et al., 2013; Yoon, 2014a) are all resources that can be utilised in the work of locating paradata in research documentation and in increasing its usefulness by creating the conditions for interpreting the paradata in dialogue with its intended functionalities and modes of production (Huvila et al., 2021c).

How useful paradata ‘in the wild’ can be identified, and what such paradata might look like, ultimately depends on the nature of the research for whose purposes the paradata is sought, and the characteristics of the obtainable research documentation. Research documentation varies enormously across

disciplines but also between different research endeavours within the same discipline and area of research. However, knowing how to mobilise the resources necessary to attain the technical and epistemic usefulness thresholds is key in identifying and using paradata in purposeful ways in many data reuse scenarios.

Being able to 'read' research documentation for paradata is a competence that is sociotechnically organised and, as underscored by Law and Lynch (1988), Niu (2009), and Kansa and Kansa (2021), a practice that can be trained. Variance in intellectual and epistemic traditions or in how methods are applied and data created or managed will always be present, but as this chapter explains there are also structures that are technically and epistemically pervasive and can be used as markers when consulting research documentation for paradata – for example, commonplace documentation types and 'genres' or recurring patterns in how paradata is recorded in these documentation types. The discussion of the ability to see, read and interact with paradata in the pursuit of multiple purposes as a competency and as a state of mind is continued in Chapters 6 and 8, where paradata literacies and paradata mindsets are discussed, respectively.

Returning to the initial discussion about differences between data and information, it can be observed that paradata crucially functions as data, in the sense that it carries the potentiality of use, but that this potentiality is realised (that is, being made informative) when it is mobilised into the practices of identifying and using paradata for reuse purposes. From this perspective, paradata emerges as an interpretable property of the connections that can be made between, within and throughout the available research documentation and the technical and epistemic resources of the party or parties collating and reading the paradata.

It does remain important to ask *how paradata can be identified* and what documentation types might be useful to collect. A complementing inquiry of equal importance might be *how paradata can be made useful* by tracing the linkages between available types of research documentation, what is known about the context from which the paradata emerged, and how the links apply to the data reuse venture at hand.

References

- Abtahi P., Ding V., Yang A. C., Bruzzese T., Romanos A. B., Murnane E. L., Follmer, S. and Landay J. A. (2020). Understanding physical practices and the role of technology in manual self-tracking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4(4). <https://doi.org/10.1145/3432236>

- Baker K. S. and Yarmey L. (2009). Data stewardship: Environmental data curation and a web-of-repositories. *International Journal of Digital Curation* 4(2), 12–27. <https://doi.org/10.2218/ijdc.v4i2.90>
- Becher T. (1989). *Academic Tribes and Territories: Intellectual Enquiry and the Cultures of Disciplines*. Milton Keynes: Society for Research into Higher Education.
- Beretta F. (2024). Semantic data for humanities and social sciences (SDHSS): An ecosystem of CIDOC CRM extensions for research data production and reuse. arXiv preprint arXiv:2402.07531. <https://doi.org/10.48550/arXiv.2402.07531>
- Berg M. (1996). Practices of reading and writing: The constitutive role of the patient record in medical work. *Sociology of Health & Illness* 18(4), 499–524. <https://doi.org/10.1111/1467-9566.ep10939100>
- Berg M. and Bowker G. (1997). The multiple bodies of the medical record: Toward a sociology of an artifact. *The Sociological Quarterly* 38(3), 513–537. <https://doi.org/10.1111/j.1533-8525.1997.tb00490.x>
- Berg M. and Goorman E. (1999). The contextual nature of medical information. *International Journal of Medical Informatics* 56(1), 51–60. [https://doi.org/10.1016/S1386-5056\(99\)00041-6](https://doi.org/10.1016/S1386-5056(99)00041-6)
- Birnholz J. P. and Bietz M. J. (2003). Data at work: Supporting sharing in science and engineering. In *Proceedings of the 2003 ACM International Conference on Supporting Group Work*, 339–348. <https://doi.org/10.1145/958160.958215>
- Borgman C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology* 63(6), 1059–1078. <https://doi.org/10.1002/asi.22634>
- Borgman C. L. (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, MA: MIT Press.
- Börjesson L. (2021). Legacy in the making: A knowledge infrastructural perspective on systems for archeological information sharing. *Open Archaeology* 7(1), 1636–1647. <https://doi.org/10.1515/opar-2020-0213>
- Börjesson L., Huvila I. and Sköld O. (2022). Information needs on research data creation. *Information Research* 27. <https://doi.org/10.47989/irisic2208>
- Börjesson L. and Sköld O. (2021). Interview dataset metadata for ‘The making and use of paradata: An interview study’. Uppsala University DiVA. Available at <https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-455730>
- Börjesson L., Sköld O., Friberg Z., Löwenborg D., Palsson G. and Huvila I. (2022b). Re-purposing excavation database content as paradata: An explorative analysis of paradata identification challenges and opportunities. *KULA: Knowledge Creation, Dissemination, and Preservation Studies* 62(3), 1–18. <https://doi.org/10.18357/kula.221>
- Börjesson L., Sköld O. and Huvila, I. (2020). Paradata in documentation standards and recommendations for digital archaeological visualisations. *Digital Culture & Society* 6(2), 191–220. <https://doi.org/10.14361/dcs-2020-0210>
- Briet S. (2006). What is documentation? In Day R. E., Martinet L. and Anghelescu H. G. B. (eds.), *What is Documentation? English Translation of the Classic French Text*. Lanham, MD: Scarecrow Press.
- Brown J. S. and Duguid P. (1996). The social life of documents. *First Monday* 1(1). <https://firstmonday.org/ojs/index.php/fm/article/download/466/387>

- Bruseker G., Carboni N. and Guillem A. (2017). Cultural heritage data management: The role of formal ontology and CIDOC CRM. In Vincent M., López-Menchero Bendicho V., Ioannides M. and Levy T. (eds.), *Heritage and Archaeology in the Digital Age*. Cham: Springer. https://doi.org/10.1007/978-3-319-65370-9_6
- Buckland M. (1991). Information as thing. *Journal of the American Society for Information Science* 42(5), 351–360. [https://doi.org/10.1002/\(SICI\)1097-4571\(199106\)42:5<351::AID-ASIS>3.0.CO;2-](https://doi.org/10.1002/(SICI)1097-4571(199106)42:5<351::AID-ASIS>3.0.CO;2-)
- Buckland M. (2018). Document theory. *Knowledge Organization* 45(5), 425–436. <https://doi.org/10.5771/0943-7444-2018-5-425>
- Carlson S. and Anderson B. (2007). What are data? The many kinds of data and their implications for data re-use. *Journal of Computer-Mediated Communication* 12(2), 635–651. <https://doi.org/10.1111/j.1083-6101.2007.00342.x>
- Ciborra C. U. and Lanzara G. F. (1994). Formative contexts and information technology: Understanding the dynamics of innovation in organizations. *Accounting, Management and Information Technologies* 4(2):61–86. [https://doi.org/10.1016/0959-8022\(94\)90005-1](https://doi.org/10.1016/0959-8022(94)90005-1)
- Cragin M. H. and Shankar K. (2006). Scientific data collections and distributed collective practice. *Computer Supported Cooperative Work* 15(2), 185–204. <https://doi.org/10.1007/s10606-006-9018-z>
- Doerr M., Ore C.-E. and Stead S. (2007). The CIDOC conceptual reference model: A new standard for knowledge sharing. In *Proceedings from the 26th International Conference on Conceptual Modeling*, 83, 51–56.
- Durrant G. B., D'Arrigo J. and Steele F. (2011). Using paradata to predict best times of contact, conditioning on household and interviewer influences. *Journal of the Royal Statistical Society Series A: Statistics in Society* 174(4), 1029–1049. <https://doi.org/10.1111/j.1467-985X.2011.00715.x>
- Edwards R., Goodwin J., O'Connor H., and Phoenix A. (2017). *Working with Paradata, Marginalia and Fieldnotes: The Centrality of By-Products of Social Research*. Northampton, MA: Edward Elgar Publishing.
- Faniel I., Kansa E., Whitcher Kansa S., Barrera-Gomez J. and Yakel E. (2013). The challenges of digging data: A study of context in archaeological data reuse. In *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries*, 295–304. New York: ACM. <https://doi.org/10.1145/2467696.2467712>
- Faniel I. and Yakel E. (2017). Practices do not make perfect: Disciplinary data sharing and reuse practices and their implications for repository data curation. In Johnston LR (ed.), *Curating Research Data, Volume One: Practical Strategies for Your Data Repository*, 103–126. Chicago, IL: ACRL.
- Faniel I. and Zimmerman A. S. (2011). Beyond the data deluge: A research agenda for large-scale data sharing and reuse. *The International Journal of Digital Curation* 1(6), 58–68. <https://doi.org/10.2218/ijdc.v6i1.172>
- Faniel I. M. and Jacobsen T. E. (2010). Reusing scientific data: How earthquake engineering researchers assess the reusability of colleagues' data. *Computer Supported Cooperative Work* 19(3), 355–375. <https://doi.org/10.1007/s10606-010-9117-8>
- Faniel I. M., Kriesberg A. and Yakel E. (2012). Data reuse and sensemaking among novice social scientists. *Proceedings of the American Society for Information Science and Technology* 49(1), 1–10. <https://doi.org/10.1002/meet.14504901068>

- Faniel I. M., Kriesberg A. and Yakel E. (2016). Social scientists' satisfaction with data reuse. *Journal of the Association for Information Science and Technology* 67(6), 1404–1416. <https://doi.org/10.1002/asi.23480>
- Fear K. and Donaldson D. R. (2012). Provenance and credibility in scientific data repositories. *Archival Science* 12(3), 319–339. <https://doi.org/10.1007/s10502-012-9172-7>
- Frank R. D., Yakel E. and Faniel I. M. (2015). Destruction/reconstruction: preservation of archaeological and zoological research data. *Archival Science* 15(2), 141–167. <https://doi.org/10.1007/s10502-014-9238-9>
- Franks, P. (2024). The crucial role of paradata in AI governance. In Duranti L. and Rogers C. (eds.), *Artificial Intelligence and Documentary Heritage*. Vancouver: InterPARES Trust AI.
- Frohmann B. (2004). *Deflating Information: From Science Studies to Documentation*. Toronto: University of Toronto Press.
- Gant S. and Reilly P. (2018). Different expressions of the same mode: A recent dialogue between archaeological and contemporary drawing practices. *Journal of Visual Art Practice* 17(1), 100–120. <https://doi.org/10.1080/14702029.2017.1384974>
- Geiger R. S. and Ribes D. (2011). Trace ethnography: Following coordination through documentary practices. In *Proceedings from the 44th Hawaii International Conference on System Sciences*. New York: IEEE. <https://doi.org/10.1109/HICSS.2011.455>
- Gitelman L. (2013). *'Raw Data' is an Oxymoron*. Cambridge, MA: MIT Press.
- Gregory K. M., Cousijn H., Groth P., Scharnhorst A. and Wyatt S. (2019). Understanding data search as a socio-technical practice. *Journal of Information Science* 46(4), 459–475. <https://doi.org/10.1177/016555151983718>
- Gunnarsson F. (2020). Digitalisation and its impact on archaeological knowledge production. In Hansson J. and Svensson J. (eds.), *Doing Digital Humanities: Concepts, Approaches, Cases*. Växjö: Linnaeus University Press, 27–44.
- Harper R. H. R. (1998). *Inside the IMF: An Ethnography of Documents, Technology, and Organisational Action*. San Diego, CA: Academic Press.
- Harvey F. and Chrisman N. (1998). Boundary objects and the social construction of GIS technology. *Environment and Planning A: Economy and Space* 30(9), 1683–1694. <https://doi.org/10.1068/a30168>
- Hilgartner S. and Brandt-Rauf S. I. (1994). Data access, ownership, and control: Toward empirical studies of access practices. *Knowledge* 15(4), 355–372. <https://doi.org/10.1177/107554709401500401>
- Hjørland B. (2000). Documents, memory institutions and information science. *Journal of Documentation* 56(1), 27–41. <https://doi.org/10.1108/EUM0000000007107>
- Hjørland B. (2018). Data (with big data and database semantics). *Knowledge Organization* 45(8), 685–708. <https://doi.org/10.5771/0943-7444-2018-8-685>
- Hjørland B. (2023). Description: Its meaning, epistemology, and use with emphasis on information science. *Journal of the Association for Information Science and Technology* 74(13), 1532–1549. <https://doi.org/10.1002/asi.24834>
- Hodges J. A. (2021). Forensically reconstructing biomedical maintenance labor: PDF metadata under the epistemic conditions of COVID-19. *Journal of the Association for Information Science and Technology* 72(11), 1400–1414. <https://doi.org/10.1002/asi.24484>

- Huggett J. (2012). Promise and paradox: Accessing open data in archaeology. In Mills C., Pidd M. and Ward E. (eds.), *Proceedings of the Digital Humanities Congress 2012: Studies in the Digital Humanities*. Sheffield: HRI Online Publications. <http://dx.doi.org/10.17613/qc5d-9x46>
- Huggett J. (2023). Extending discourse analysis in archaeology: A multimodal approach. In Gonzalez-Perez C., Martin-Rodilla P. and Pereira-Fariña M. (eds.), *Discourse and Argumentation in Archaeology: Conceptual and Computational Approaches*. Cham: Springer. https://doi.org/10.1007/978-3-031-37156-1_8
- Huvila, I. (2011). The politics of boundary objects: Hegemonic interventions and the making of a document. *Journal of the American Society for Information Science and Technology* 62(12), 2528–2539. <https://doi.org/10.1002/asi.21639>
- Huvila I. (2014). Archaeologists and their information sources. In Huvila I. (ed.), *Perspectives to Archaeological Information in the Digital Society*. Uppsala: Uppsala University, Department of ALM, 25–54.
- Huvila I. (2020). Information-making-related information needs and the credibility of information. *Information Research*, 25(4). <https://doi.org/10.47989/irisic2002>
- Huvila I. (2022). Improving the usefulness of research data with better paradata. *Open Information Science* 6(1), 28–48. <https://doi.org/10.1515/opis-2022-0129>
- Huvila I., Andersson L., Fulton C., Haider J. and Harviainen J. T. (2023a). Managing information gaps and non-information. *Proceedings of the Association for Information Science and Technology* 60(1), 793–798. <https://doi.org/10.1002/pra2.863>
- Huvila I., Andersson L., and Sköld O. (2022a). Citing methods literature: Citations to field manuals as paradata on archaeological fieldwork. *Information Research* 27(3). <https://doi.org/10.47989/irpaper941>
- Huvila I., Börjesson L., and Sköld O. (2022b). Archaeological information-making activities according to field reports. *Library & Information Science Research* 44(3). <https://doi.org/10.1016/j.lisr.2022.101171>
- Huvila I., Greenberg J., Sköld O., Thomer A., Trace C. and Zhao X. (2021a). Documenting information processes and practices: Paradata, provenance metadata, life-cycles and pipelines. *Proceedings of the Association for Information Science and Technology* 58(1), 604–609. <https://doi.org/10.1002/pra2.509>
- Huvila I. and Sköld O. (2023). A fieldwork manual as a regulatory device: Instructing, prescribing and describing documentation work. *Journal of Information Science*. <https://doi.org/10.1177/01655515231203>
- Huvila I., Sköld O., and Andersson L. (2023b). Knowing-in-practice, its traces and ingredients. In Cozza M. and Gherardi S. (eds.), *The Posthumanist Epistemology of Practice Theory*. Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-031-42276-8_2
- Huvila I., Sköld O. and Börjesson L. (2021). Documenting information making in archaeological field reports. *Journal of Documentation* 77(5), 1107–1127. <https://doi.org/10.1108/JD-11-2020-0188>
- Kansa E. and Kansa S. W. (2021). Digital data and data literacy in archaeology now and in the new decade. *Advances in Archaeological Practice* 9(1), 81–85. <https://doi.org/10.1017/aap.2020.55>
- Kim Y. and Yoon A. (2017). Scientists' data reuse behaviors: A multilevel analysis. *Journal of the Association for Information Science and Technology* 68(12), 2709–2719. <http://doi.org/10.3886/E100404V1>

- Kintigh K. (2006). The promise and challenge of archaeological data integration. *American Antiquity* 71(3), 567–578. <http://doi.org/10.2307/40035365>
- Kirschenbaum M. G. (2008). *Mechanisms: New Media and the Forensic Imagination*. Cambridge, MA: MIT Press.
- Knorr-Cetina K. (1999). *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, MA: Harvard University Press.
- Koesten L., Simperl E., Blount T., Kacprzak E. and Tennison J. (2020). Everything you always wanted to know about a dataset: Studies in data summarisation. *International Journal of Human-Computer Studies* 135. <https://doi.org/10.1016/j.ijhcs.2019.10.004>
- Latour B. (1987). *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press.
- Latour B. (1999). Circulating reference: Sampling the soil in the Amazon Forest. In *Pandora's Hope: Essays on the Reality of Science Studies*. Cambridge, MA: Harvard University Press, 25–79
- Latour B. and Woolgar S. (1979). *Laboratory Life: The Social Construction of Scientific Facts*. Beverly Hills, CA: Sage.
- Law J. (1993). *Organising Modernity: Social Ordering and Social Theory*. Hoboken, NJ: Wiley.
- Law J. and Lynch M. (1988). Lists, field guides, and the descriptive organization of seeing: Birdwatching as an exemplary observational activity. *Human Studies* 11(2/3), 271–303.
- Leonelli S. and Tempini N. (eds.) (2020). *Data Journeys in the Sciences*. Cham: Springer.
- Levy D. M. (2001). *Scrolling Forward: Making Sense of Documents in the Digital Age*. New York: Arcade Pub.
- Lund N. W. (2009). Document theory. *Annual Review of Information Science and Technology* 43, 399–432. <https://doi.org/10.1002/aris.2009.1440430116>
- Lund N. W. (2010). Document, text and medium: Concepts, theories and disciplines. *Journal of Documentation* 66(5), 734–749. <https://doi.org/10.1108/00220411011011066817>
- Maron D. and Feinberg M. (2018). What does it mean to adopt a metadata standard? A case study of Omeka and the Dublin Core. *Journal of Documentation*, 74(4), 674–691. <https://doi.org/10.1108/JD-06-2017-0095>
- Mayernik M. S. (2016). Research data and metadata curation as institutional issues. *Journal of the Association for Information Science and Technology* 67(4), 973–993. <https://doi.org/10.1002/asi.23425>
- Nadim T. (2021). The datafication of nature: Data formations and new scales in natural history. *Journal of the Royal Anthropological Institute* 27(1), 62–75. <https://doi.org/10.1111/1467-9655.13480>
- Niu J. (2009). Overcoming inadequate documentation. *Proceedings of the American Society for Information Science and Technology* 46(1), 1–14. <https://doi.org/10.1002/meet.2009.145046024>
- Oliver G., Cranefield J., Lilley S. and Lewellen M. J. (2024). Understanding data culture/s: Influences, activities, and initiatives: An Annual Review of Information Science and Technology (ARIST) paper. *Journal of the Association for Information Science and Technology* 75(3), 201–214. <https://doi.org/10.1002/asi.24737>

- Olson H. A. (2002). *Power to Name: Locating the Subject Representation in Libraries*. Boston: Kluwer.
- Pardo T. A., Cresswell A. M., Thompson F. and Zhang J. (2006). Knowledge sharing in cross-boundary information system development in the public sector. *Information Technology and Management* 7(4), 293–313. <https://doi.org/10.1007/s10799-006-0278-6>
- Pasquetto I. V., Borgman C. L. and Wofford M. F. (2019). Uses and reuses of scientific data: The data creators' advantage. *Harvard Data Science Review* 1(2). <https://doi.org/10.1162/99608f92.fc14bf2d>
- Pickering A. (1995). *The Mangle of Practice: Time, Agency, and Science*. Chicago: University of Chicago Press.
- Pinch T. J. and Bijker W. E. (1984). The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other. *Social Studies of Science* 14(3), 399–441. <https://doi.org/10.1177/0306312840140030>
- Plantin J.-C. and Thomer A. (2023). Platforms, programmability, and precarity: The platformization of research repositories in academic libraries. *New Media & Society*. <https://doi.org/10.1177/146144482311767>
- Richards J. D., Jakobsson U., Novák D., Štular B. and Wright, H. (2021). Digital archiving in archaeology: The state of the art. *Internet Archaeology* 58. <https://doi.org/10.11141/ia.58.23>
- Rolland B. and Lee C. P. (2013). Beyond trust and reliability: Reusing data in collaborative cancer epidemiology research. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*. New York: ACM, 435–444. <https://doi.org/10.1145/2441776.2441826>
- Rowley J. (2007). The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of Information Science* 33(2), 163–180. <https://doi.org/10.1177/0165551506070706>
- Shankar K. (2007). Order from chaos: The poetics and pragmatics of scientific record-keeping. *Journal of the American Society for Information Science and Technology* 58, 1457–1466. <https://doi.org/10.1002/asi.20625>
- Shankar K. (2009). Ambiguity and legitimate peripheral participation in the creation of scientific documents. *Journal of Documentation*, 65(1), 151–165. <https://doi.org/10.1108/00220410910926167>
- Shepherd N. G. and Rudd J. M. (2014). The influence of context on the strategic decision-making process: A review of the literature. *International Journal of Management Reviews* 16(3), 340–364. <https://doi.org/10.1111/ijmr.12023>
- Sköld O. (2017). Getting-to-know: Inquiries, sources, methods, and the production of knowledge on a videogame wiki. *Journal of Documentation* 73(6), 1299–1321. <https://doi.org/10.1108/JD-11-2016-0145>
- Sköld O. (2018). *Documenting Videogame Communities: A Study of Community Production of Information in Social-media Environments and its Implications for Videogame Preservation*. PhD dissertation. Uppsala: Department of ALM, Uppsala University.
- Sköld O., Börjesson L. and Huvila I. (2022). Interrogating Paradata. *Information Research*, 27. <https://doi.org/10.47989/colis2206>

- Spedding P. and Tankard P. (eds.) (2021). *Marginal Notes: Social Reading and the Literal Margins*. Basingstoke: Palgrave Macmillan.
- Steinkuehler C. and Duncan S. (2008). Scientific habits of mind in virtual worlds. *Journal of Science Education and Technology* 17(6), 530–543. <https://doi.org/10.1007/s10956-008-9120-8>
- The London Charter Organization (2009) *The London Charter for the Computer-Based Visualisation of Cultural Heritage*. Available at <https://londoncharter.org> (accessed 12 April 2024).
- Trace C. B. (2011). Documenting work and working documents: Perspectives from workplace studies, CSCW, and genre studies. In Sprague, R. H. Jr. (ed.), *Proceedings from the 44th Hawaii International Conference on System Sciences*. New York: IEEE. <https://doi.org/10.1109/HICSS.2011.170>
- Trace C. B. and Zhang Y. (2020). The quantified-self archive: Documenting lives through self-tracking data. *Journal of Documentation* 76(1), 290–316. <https://doi.org/10.1108/JD-04-2019-0064>
- Ullah I. I. T. (2015). Integrating older survey data into modern research paradigms: Identifying and correcting spatial error in ‘legacy’ datasets. *Advances in Archaeological Practice* 3(4), 331–350. <https://doi.org/10.7183/2326-3768.3.4.331>
- Van House N. A. (2002a) Digital libraries and practices of trust: Networked biodiversity information. *Social Epistemology* 16(1), 99–114. <https://doi.org/10.1080/02691720210132833>
- Van House N. A. (2002b). Trust and epistemic communities in biodiversity data sharing. In *Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries*, 231–239. <https://doi.org/10.1145/544220.544270>
- Voss, B. L. (2012). Curation as research: A case study in orphaned and under-reported archaeological collections. *Archaeological Dialogues* 19(2), 145–169. <https://doi.org/10.1017/S1380203812000219>
- Wallis J. C., Rolando E. and Borgman C. L. (2013). If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PLoS ONE* 8(7). <https://doi.org/10.1371/journal.pone.0067332>
- Warmelink H. (2013). *Online Gaming and Playful Organization*. New York: Routledge.
- Weber N. M., Baker K. S., Thomer A. K., Chao T. C. and Palmer C. L. (2012). Value and context in data use: Domain analysis revisited. *Proceedings of the American Society for Information Science and Technology* 49(1):1–10. <https://doi.org/10.1007/s10502>
- West B. T. and Sinibaldi J. (2013). The quality of paradata: A literature review. In Kreuter F (ed.), *Improving Surveys with Paradata: Analytic Uses of Process Information*. Hoboken, NJ: Wiley. <https://doi.org/10.1002/9781118596869.ch14>
- Yakel E., Faniel I., Kriesberg A. and Yoon A. (2013). Trust in digital repositories. *The International Journal of Digital Curation* 8(1), 1–14. <https://doi.org/10.2218/ijdc.v8i1.251>
- Yoon A. (2014a). End users’ trust in data repositories: Definition and influences on trust development. *Archival Science* 14(1), 17–34.

- Yoon A. (2014b). 'Making a square fit into a circle': Researchers' experiences reusing qualitative data. *Proceedings of the American Society for Information Science and Technology* 51(1), 1–4. <https://doi.org/10.1002/meet.2014.14505101140>
- Zimmerman A. S. (2008). New knowledge from old data: The role of standards in the sharing and reuse of ecological data. *Science, Technology, & Human Values* 33(5), 631–652. <https://doi.org/10.1177/0162243907306704>